POZNAŃ
UNIVERSITY
OF LIFE SCIENCES

INSECT-IMP

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

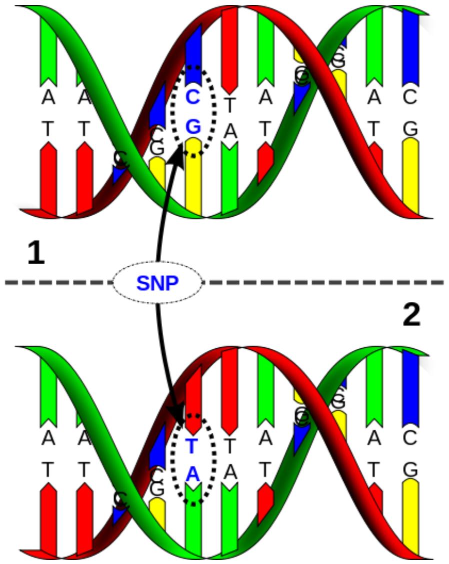
Funded by
the European Union

So I have the genotypes, what next?

Ewa Sell-Kubiak

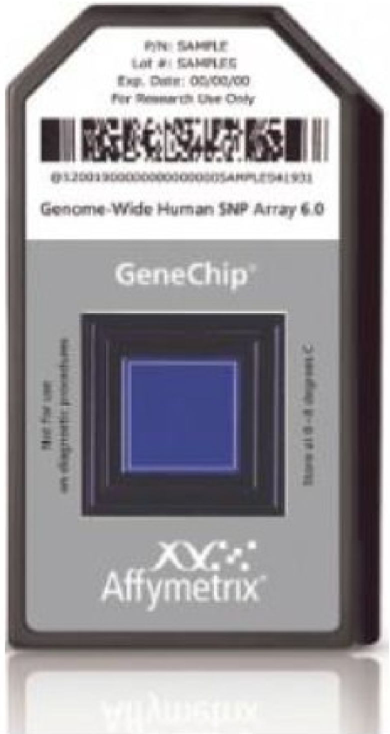
Department of Genetics and Animal Breeding
Faculty of Veterinary Medicine and Animal Science

SINGLE-NUCLEOTIDE POLYMORPHISM - SNP



Represents 90% of the variability of the entire genome

SNP-CHIPS



Affymetrix®

illumina®

SNP-CHIPS

Name	**#SNPs**
Axiom Porcine660K	> 660,000
PorcineSNP60 BeadChip	> 60,000
GeneSeek Genomic Profiler	> 10,000
Individual	??

SNP-CHIPS FOR INSECTS

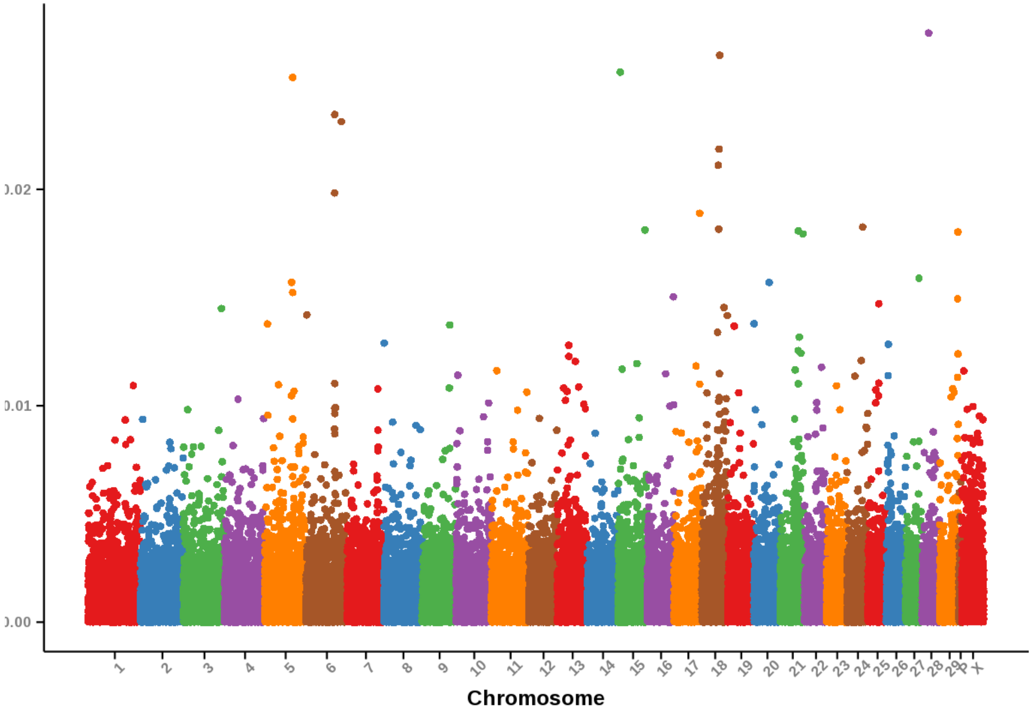
Specie	Name	#SNPs
Honeybee	**Affymetrix Axiom Apis_mellifera 660K**	> 660,000
Honeybee	**Illumina BeeHD BeadChip**	> 7,000
Mosquitoes	**Affymetrix Axiom Anopheles gambiae**	> 760,000
Silkworm	**Affymetrix Axiom Bombyx_mori_1.0**	~575,000
BSF	??	

Brought to you by ChatGPT4.0

SO I HAVE THE GENOTYPES, WHAT NEXT?

So I have the genotypes, what next?

Let’s run a Genome-wide association study!

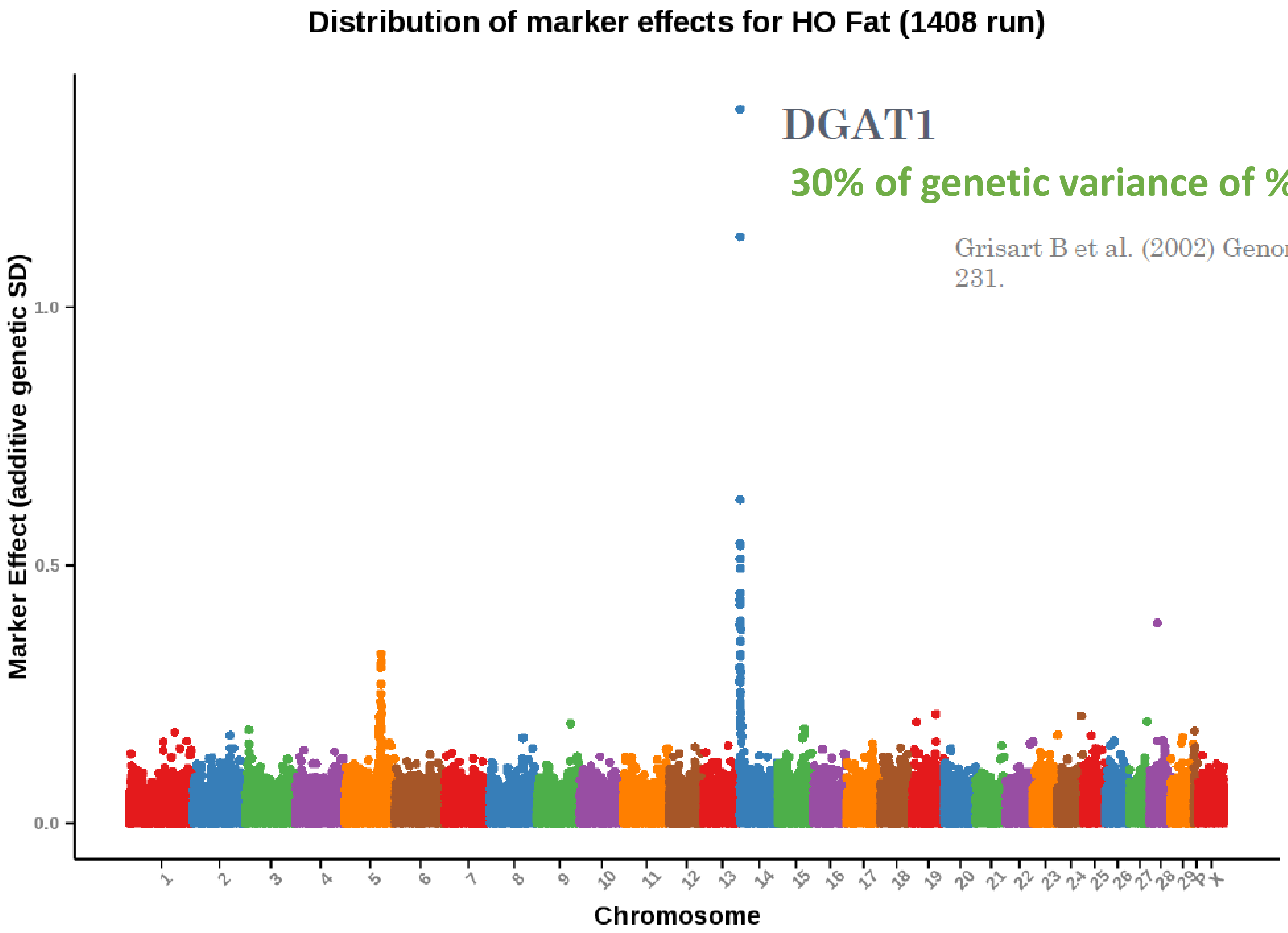


WHEN RUN GWAS?

- **To study a genetic background of a new, unexplored trait**
- **The studied traits is heritable ($h^2 > 0$)**
- **No potential candidate genes are known**

- ***A perfect result – one major gene and some of a small effect***

THE PERFECT RESULT



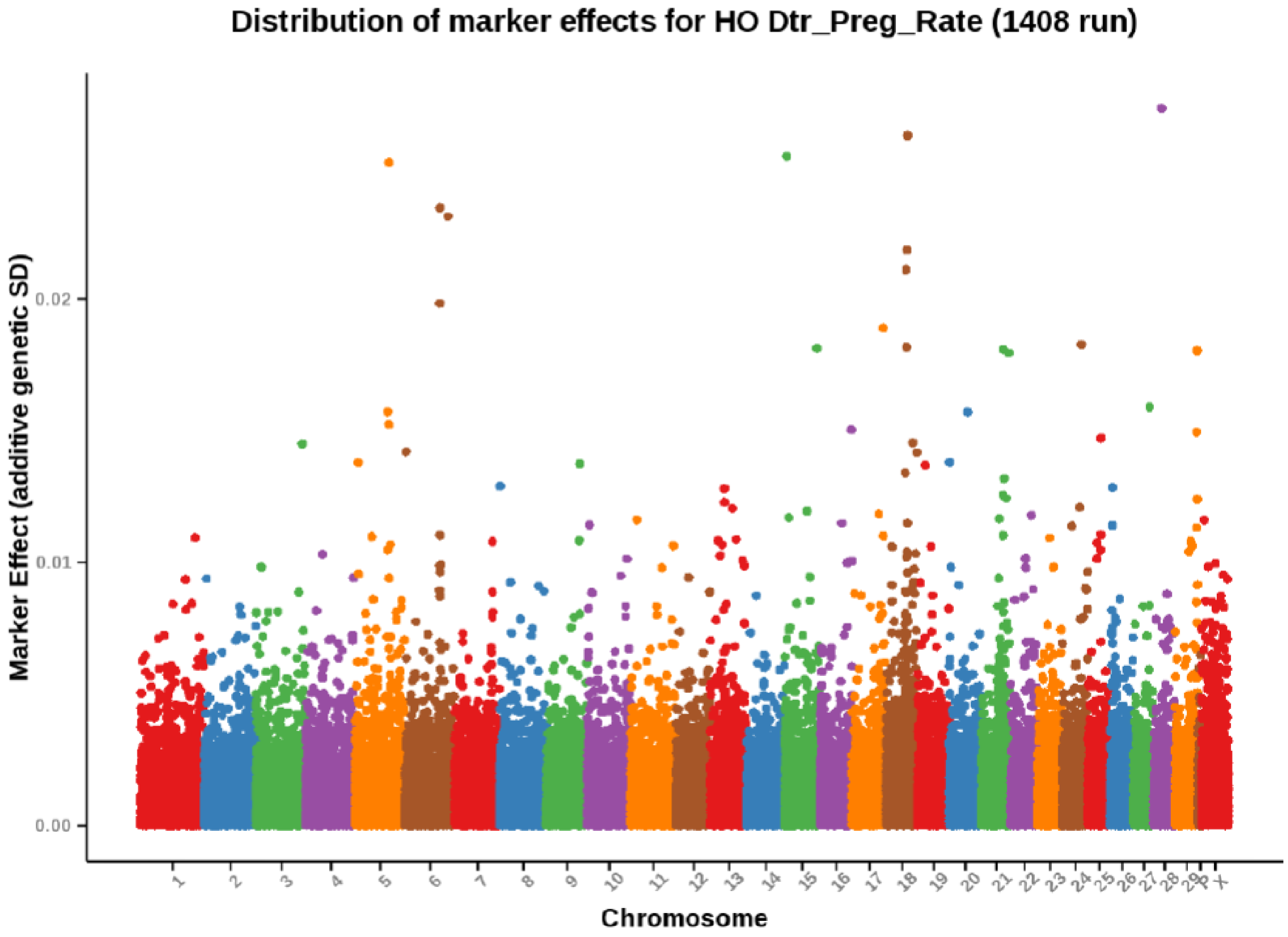
DGAT1

30% of genetic variance of %fat in milk

Grisart B et al. (2002) Genome Res 12: 222–231.

https://www.cdcb.us/Report_Data/Marker_Effects/marker_effects.cfm?Breed=HO&Trait=Fat

THE REALITY CHECK



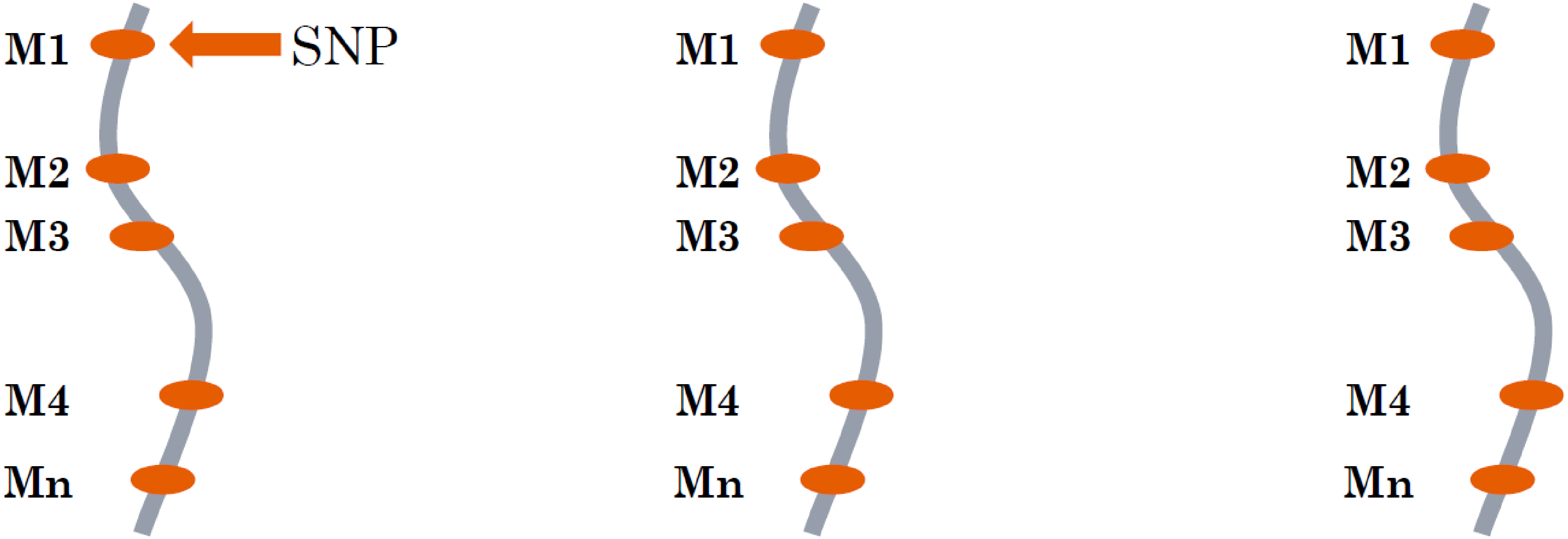
https://www.cdcb.us/Report_Data/Marker_Effects/marker_effects.cfm?Breed=HO&Trait=Dtr_Preg_Rate

How to get to ANY result?

Use statistics ☺

ASSOCIATION ANALYSIS

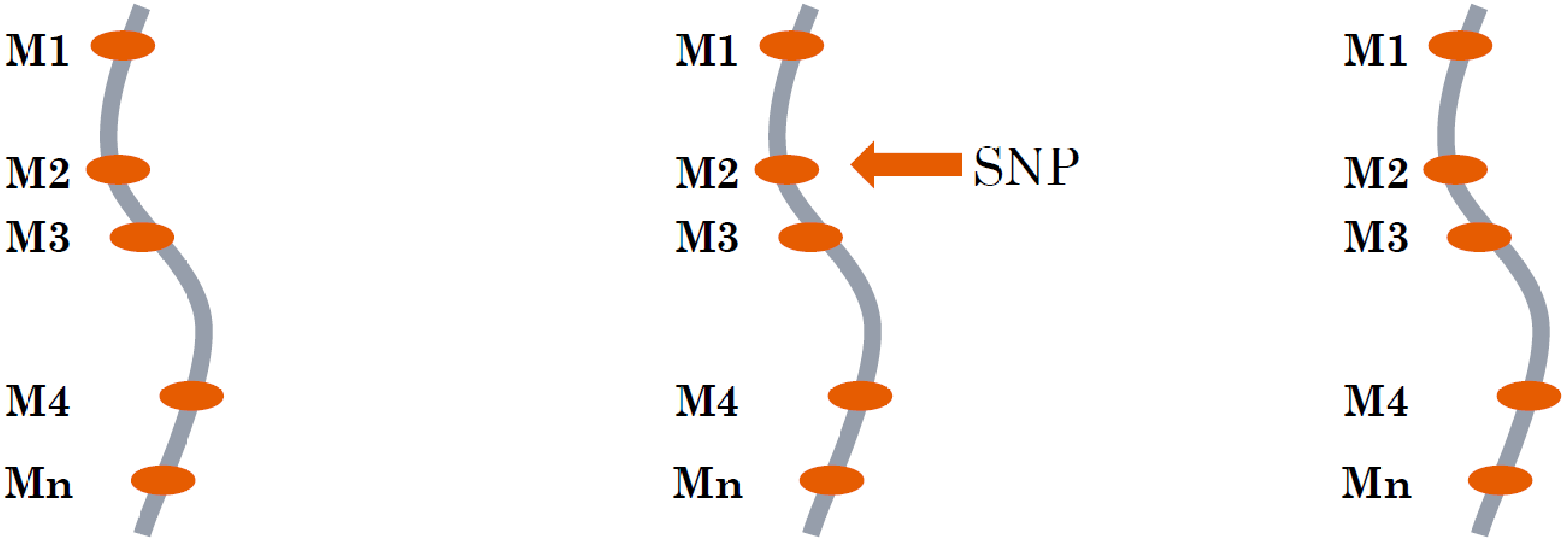
Classical – linear regression



$$\boxed{y_{ij} = \mu + M_1 + e_{ij}}$$

Association analysis

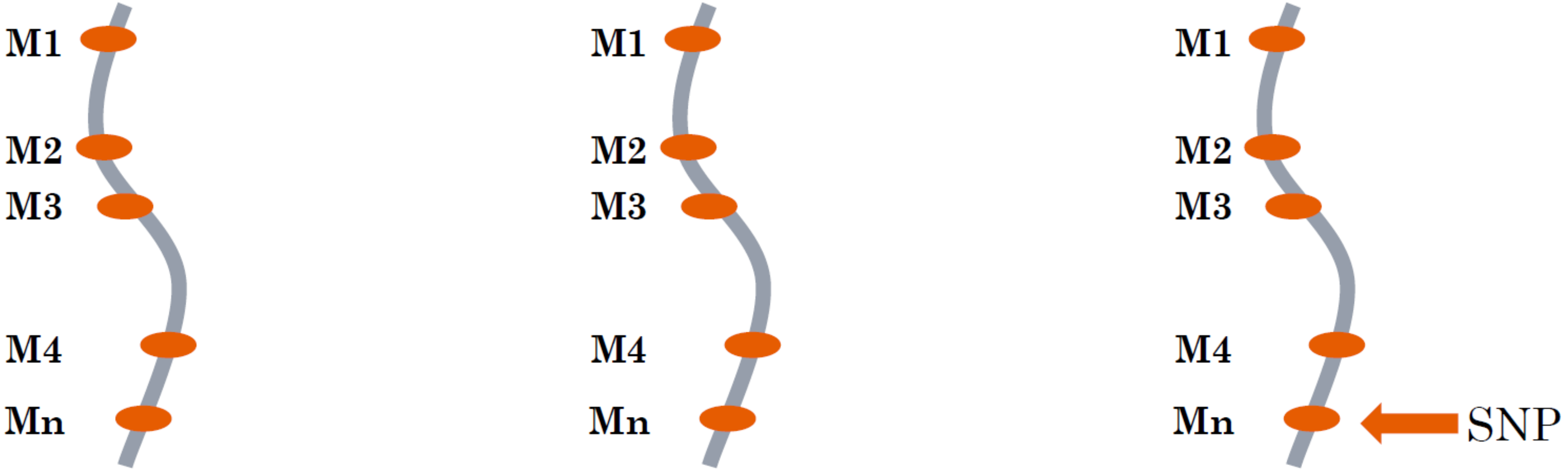
Classical – linear regression



$$\boxed{y_{ij} = \mu + M_2 + e_{ij}}$$

ASSOCIATION ANALYSIS

Classical – linear regression



$$\boxed{y_{ij} = \mu + \mathrm{M}_n + e_{ij}}$$

TYPES OF GWAS

Single-SNP with A-matrix

Single SNp with G-matrix

Classical regression (in any statistical program)

with correction for relationship between the animals

TYPES OF GWAS

Multi-SNP

Bayesian statistics (in eg. bayz)

Many SNPs is drawn at once and given a greater importance in influencing the trait of interest, the remaining SNPs are still used in the analysis to correct for relatedness in the population.

TYPES OF GWAS

Slinding windows

Only in BLUPF90 (Ignacy Miszal's program)

Windows with a specific number of SNPs slide along each chromosome. The assessment is based on area and not individual SNPs.

TYPE OF PHENOTYPIC DATA

- **Direct phenotype**

Directly measured on an individual e.g., daily gains, milk yield, number of piglets

- **Breeding value**

Solution for males based on the records of daughters (milk yield, number of piglets) or when the trait is difficult to measure

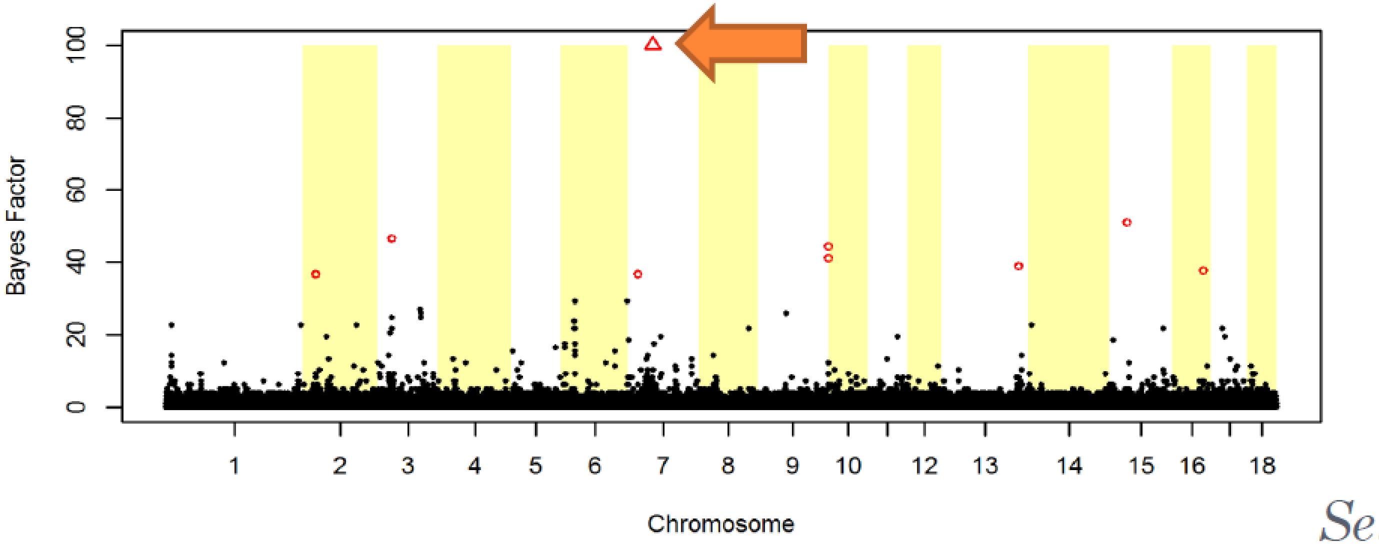
How to get to GOOD result?

Use statistics AND check quality of the data

WHAT AFFECTS THE GWAS RESULT?

- **Quality of phenotypes and genotypes**
- **Heritability of the trait**
- **Coverage of the genome by SNPs**
- **Number of genotyped animals**
- **Correctly selected statistical model (FIXED effects!!)**
- **Genomic structure of the population**

IF ALL IS DONE WELL

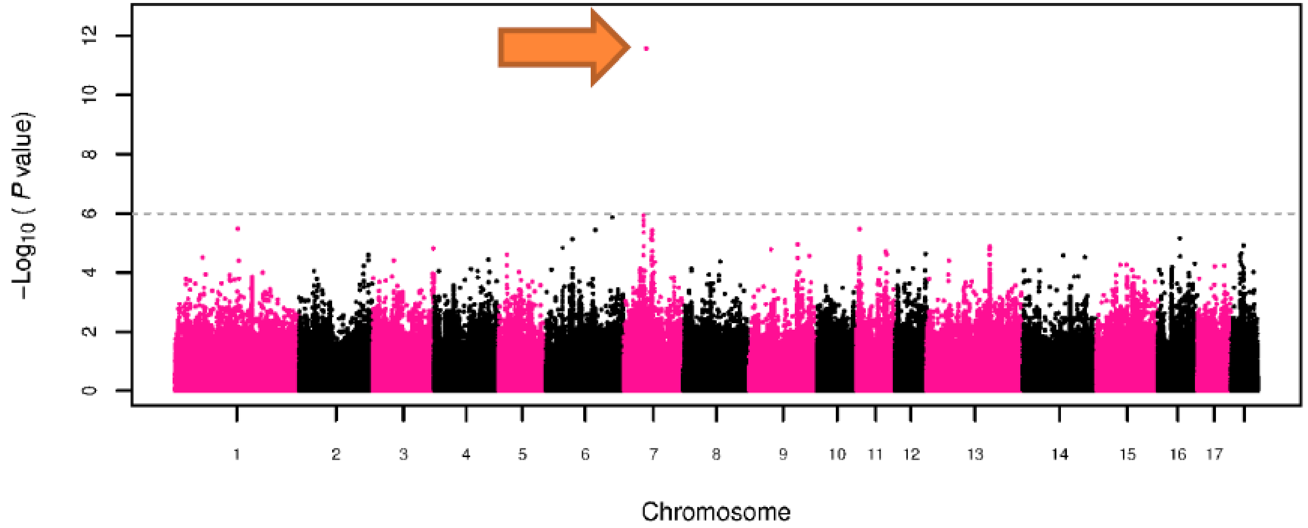


~2,000 pigs with genotype

40,969 SNPs

Method/software:
multiple SNP
GWAS in bayz

Sell-Kubiak et al. 2015, BMC Genomics



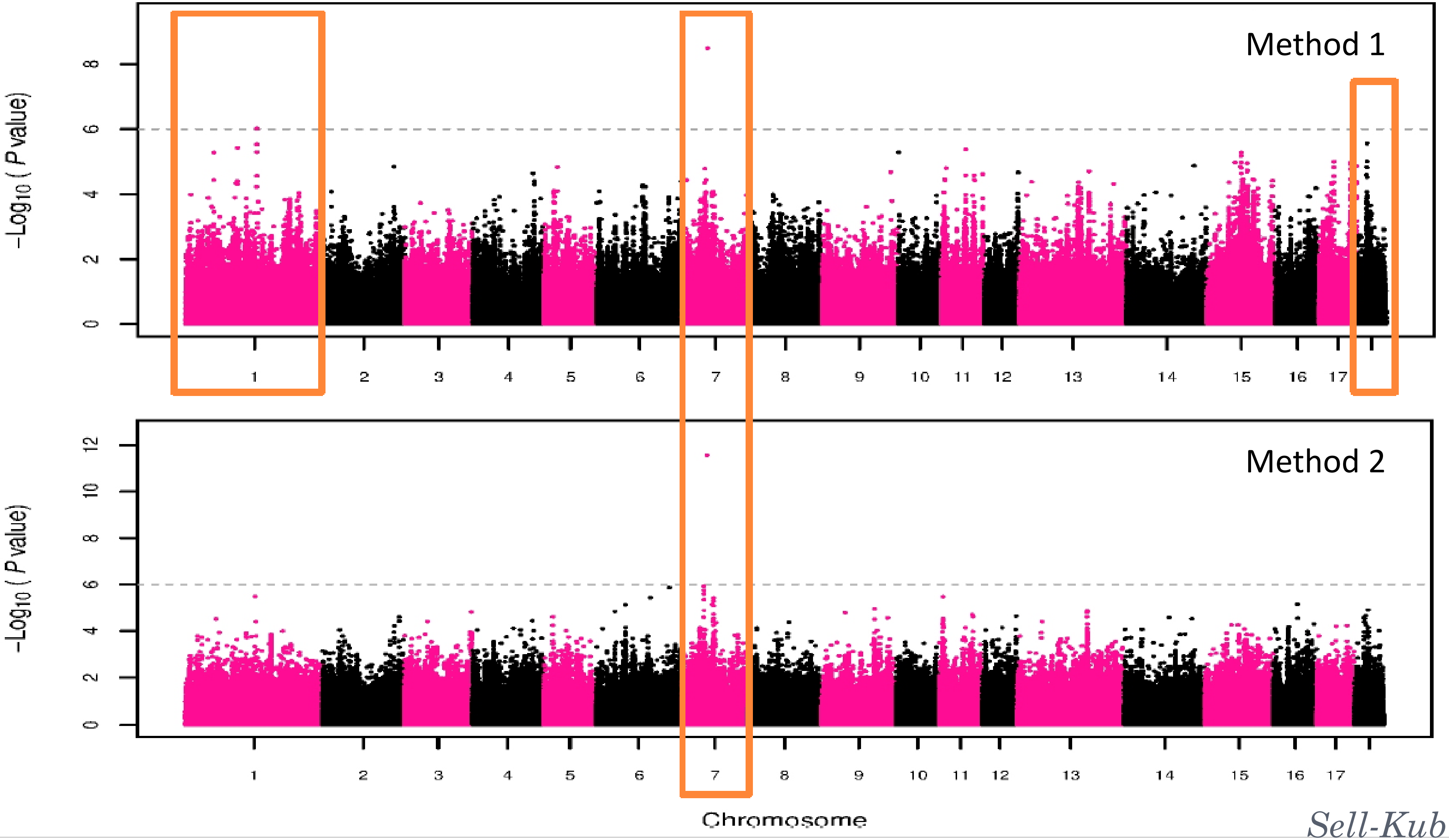
~12,000 pigs with genotype

526,000 SNPs

Method/software:
single SNP
GWAS in GCTA

Sell-Kubiak et al. 2022, GSE

FALSE POSITIVES DUE TO MODEL SELECTION

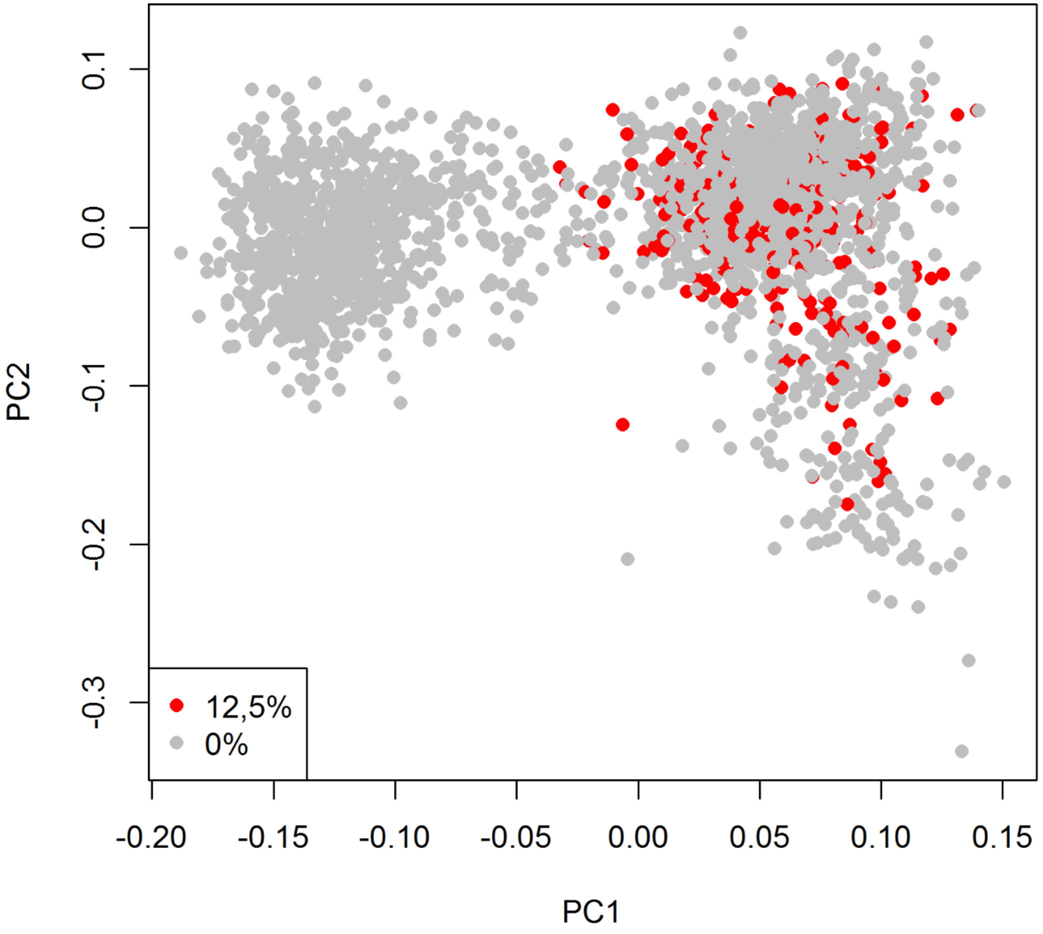


Sell-Kubiak et al., 2022

COMPLETELY INCORRECT RESULTS DUE TO IGNORED GENOMIC STRUCTURE OF THE POPULATION

Level of genomic variability in the population of Large White pigs from TopigsNorsvin

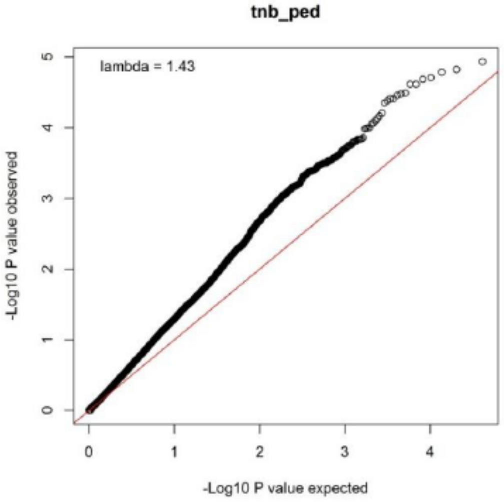
Sell-Kubiak, 2015

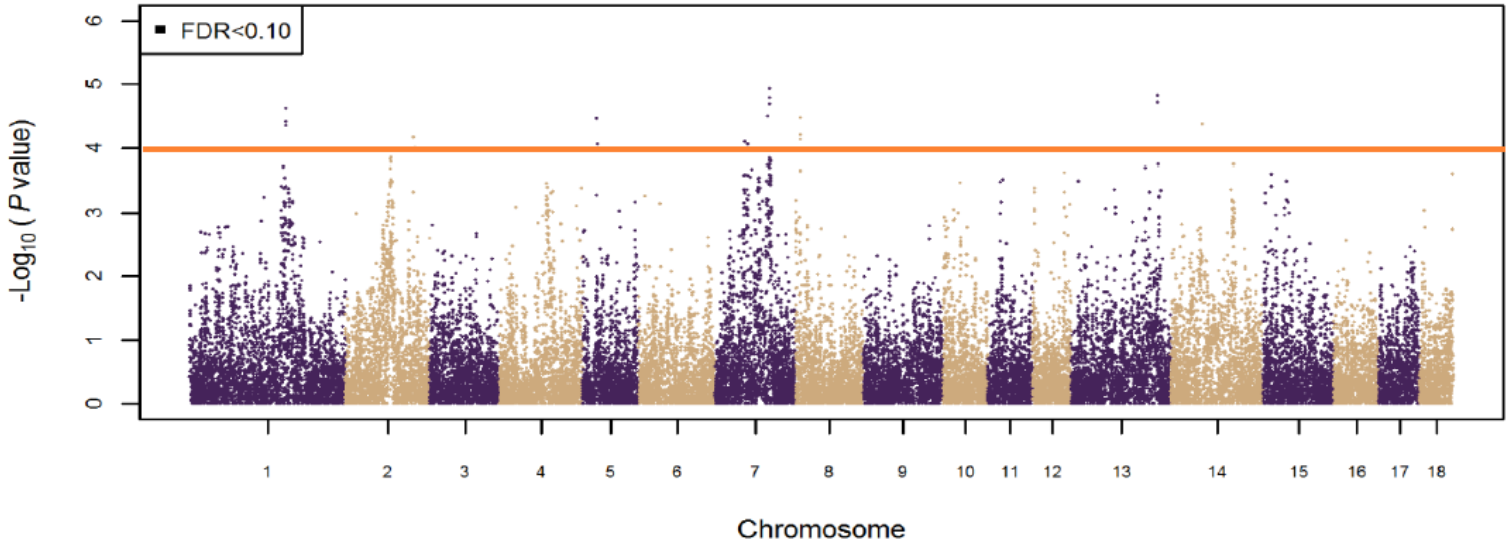


GENOMIC STRUCTURE OF POPULATION

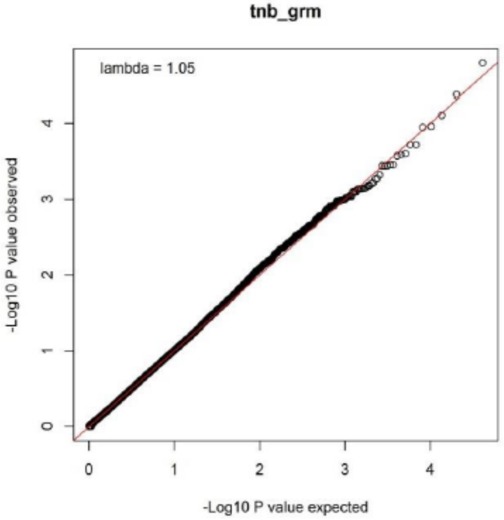
SELL-KUBIAK, 2015

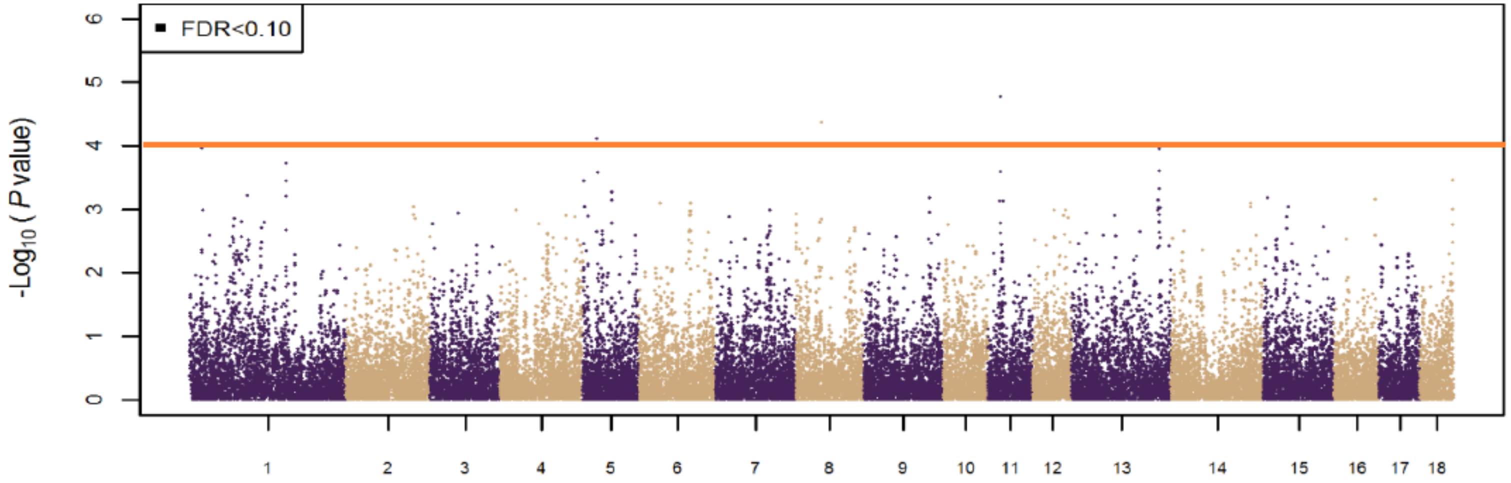
Single-SNP with A-matrix





Single-SNP with G-matrix

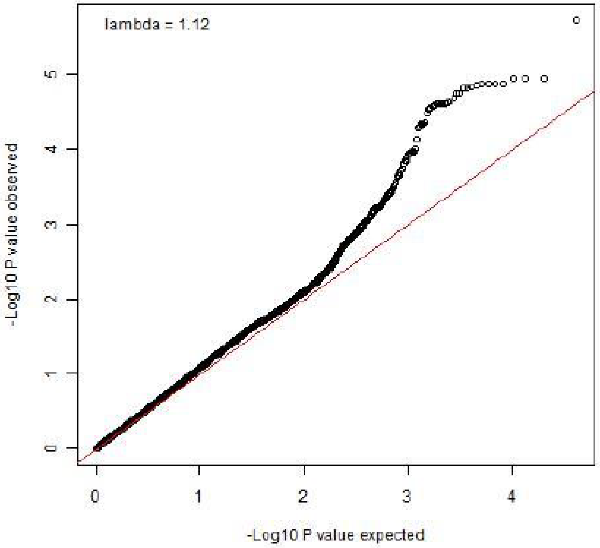


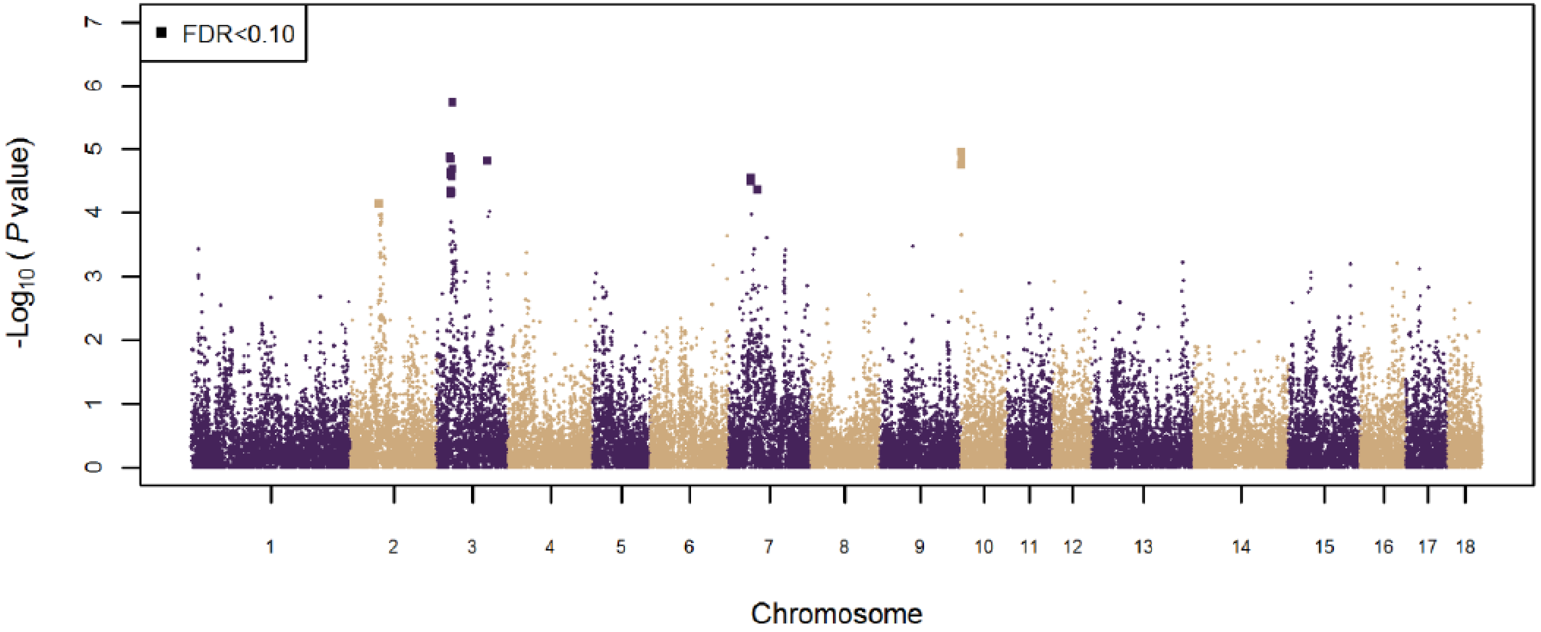


GENOMIC STRUCTURE OF POPULATION

SELL-KUBIAK, 2015

Single-SNP with A-matrix with correct population structure



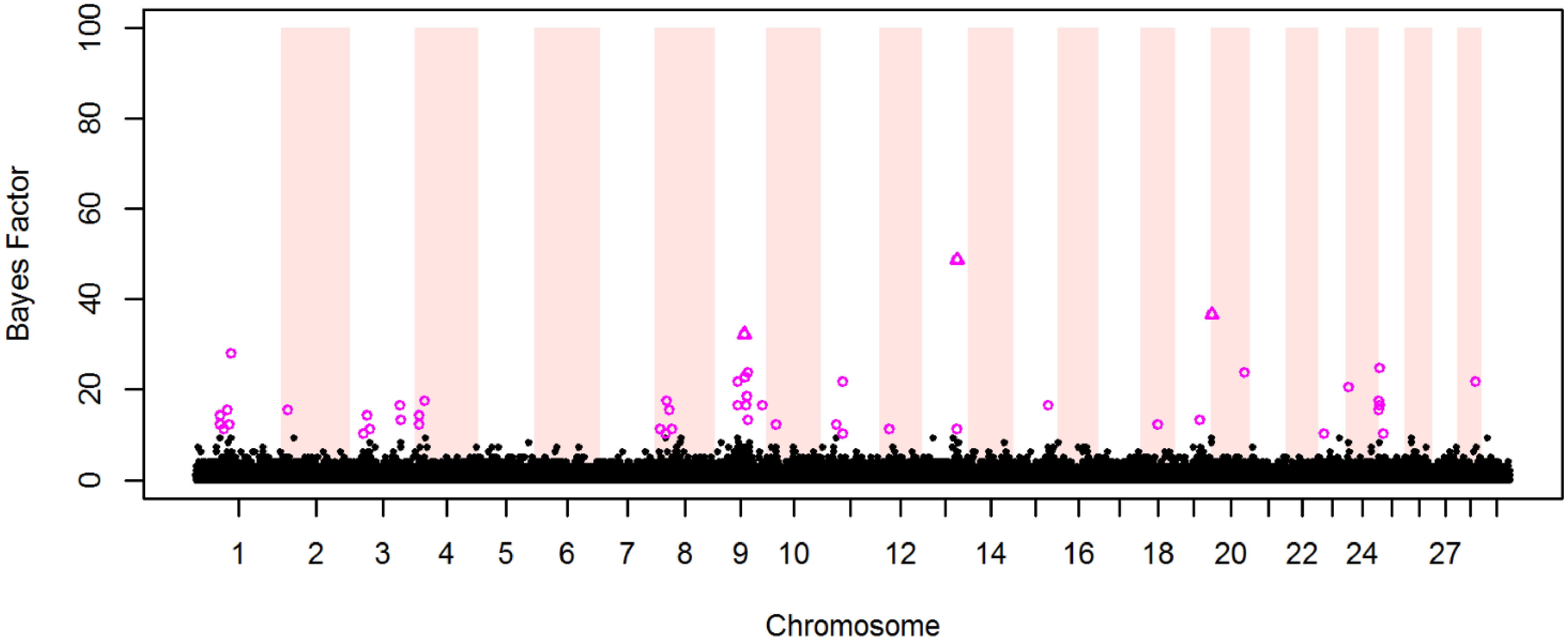


GWAS IS JUST THE BEGINING… POST-GWAS ANALYSIS

- Determining the location of the most important SNPs in the genome (chromosome + position in bp, i.e. base pairs)
- Searching available databases with a reference genome (e.g. BioMART) to determine whether the SNP is in a coding position (exon), non-coding (intron) or regulatory
- Verifying the biological significance of the gene in which (or next to which) the SNP is located e.g. in GeneCards or GO Terms analysis, if we have many candidate genes
- ***Many tools and many options…***

POST-GWAS ANALYSIS EXAMPLE

Multi-SNP GWAS for methan emission in cattle ***Pszczoła et al., 2018***



POST-GWAS ANALYSIS EXAMPLE

Multi-SNP GWAS for methan emission in cattle *Pszczoła et al., 2018*

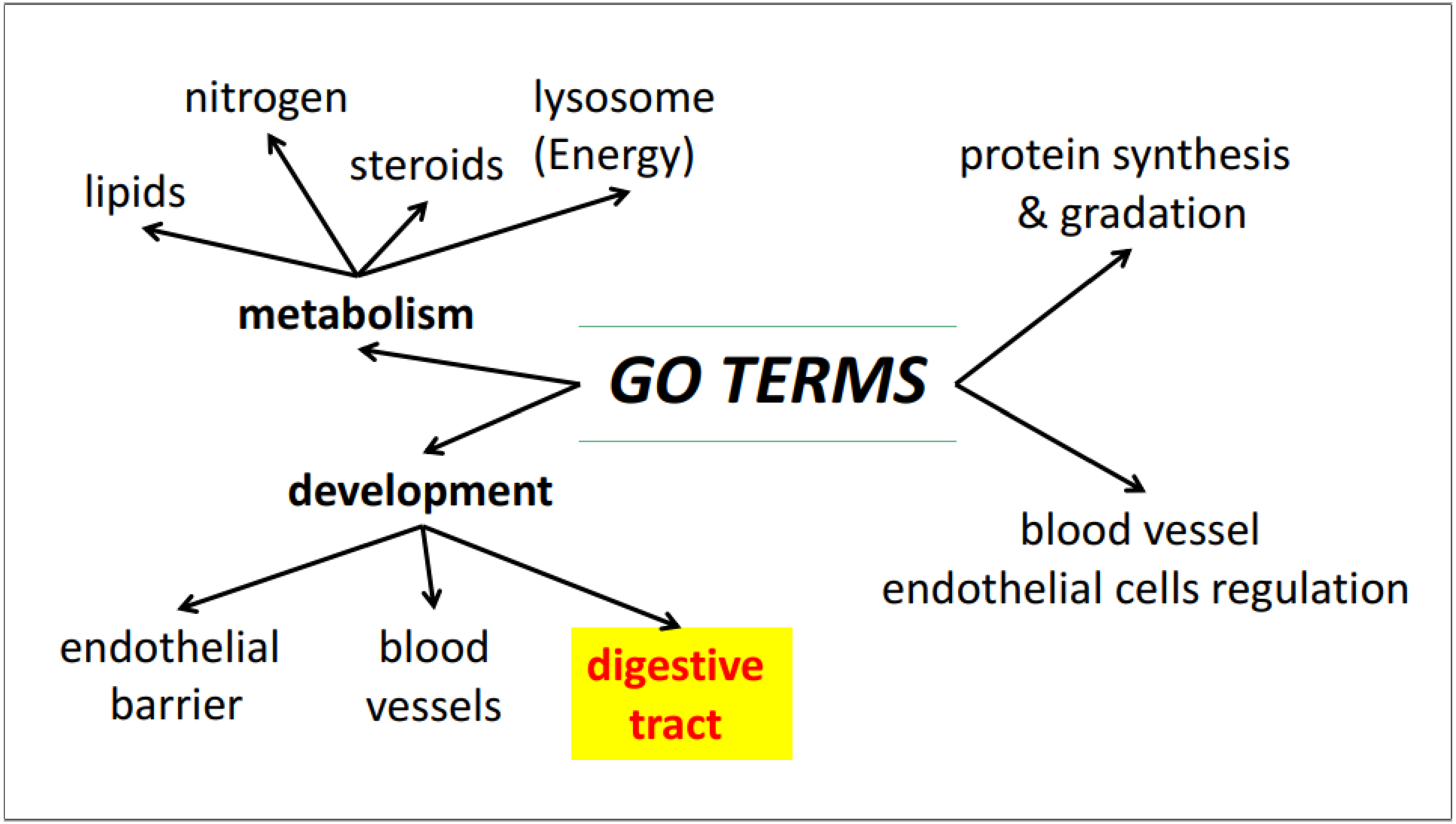
130 candidate genes

46
known & non-ambiguous genes

*BIOMART (Ensembl Bos Taurus UMD 3.1)

POST-GWAS ANALYSIS EXAMPLE

Multi-SNP GWAS for methan emission in cattle *Pszczoła et al., 2018*



GO TERMS

metabolism

lipids

nitrogen

steroids

lysosome (Energy)

development

endothelial barrier

blood vessels

digestive tract

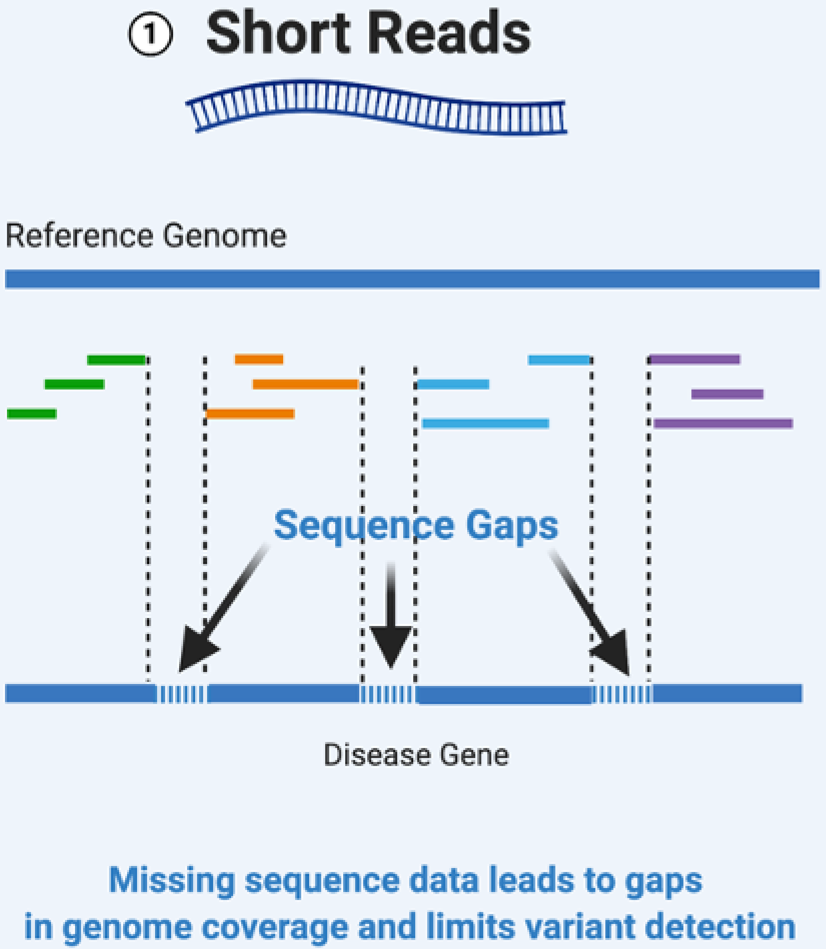
protein synthesis & gradation

blood vessel endothelial cells regulation

What if 660K SNPs is not enough for me?

Go for a whole-genome sequence!

TYPES OF SEQUENCE DATA



New generation sequencing
NGS
Illumina
Thermo Fisher Scientific

Classical and best-known approach

Great genome coverage

Allows to do everything:
GWAS
Imputations
Case-Contol studies

Hudsonalpha.org

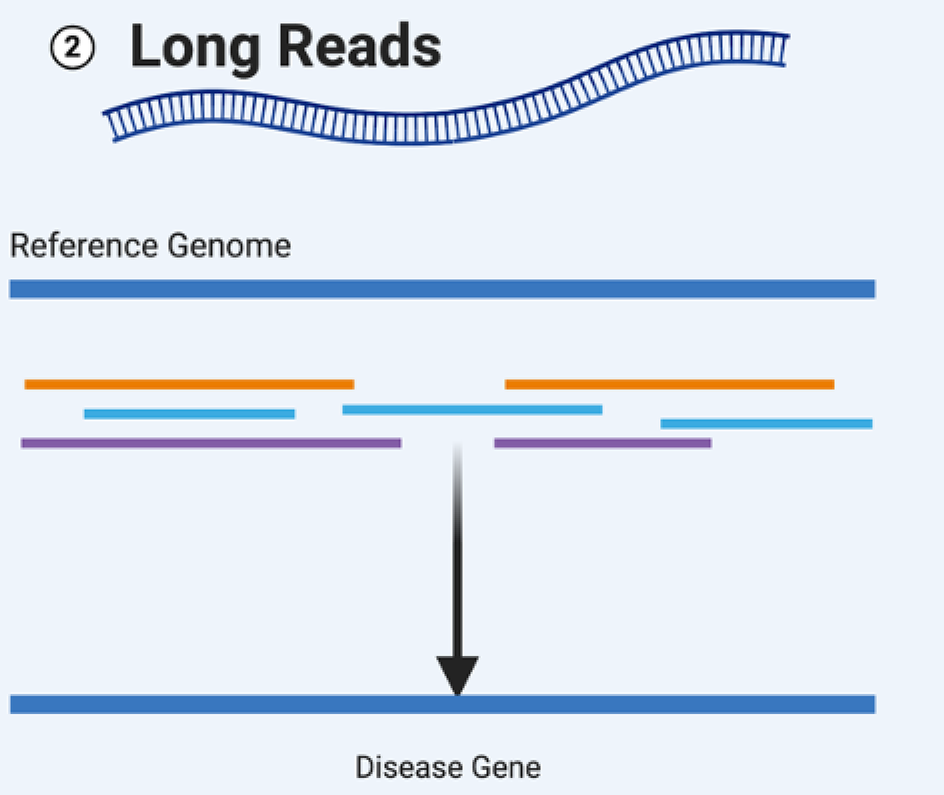
TYPES OF SEQUENCE DATA

New generation sequencing
NGS
Oxford Nanopore Technologies (ONT)

Gives more possibilities
Pan-genomes
DNA-methylation

Perfect genome coverage

Expensive



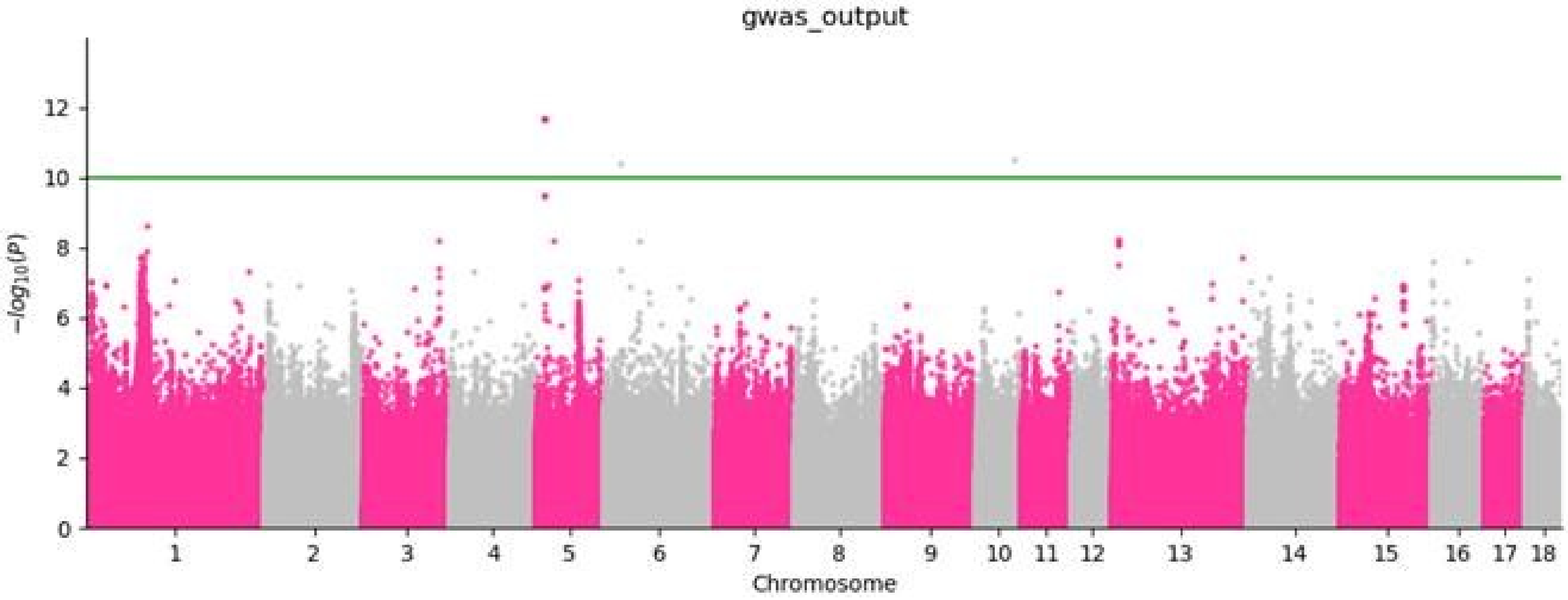
Long reads map uniquely and span large variants providing comprehensive variant detection

Hudsonalpha.org

GWAS WITH SHORT READS

CIELEŃ ET AL., TO BE SUBMITTED

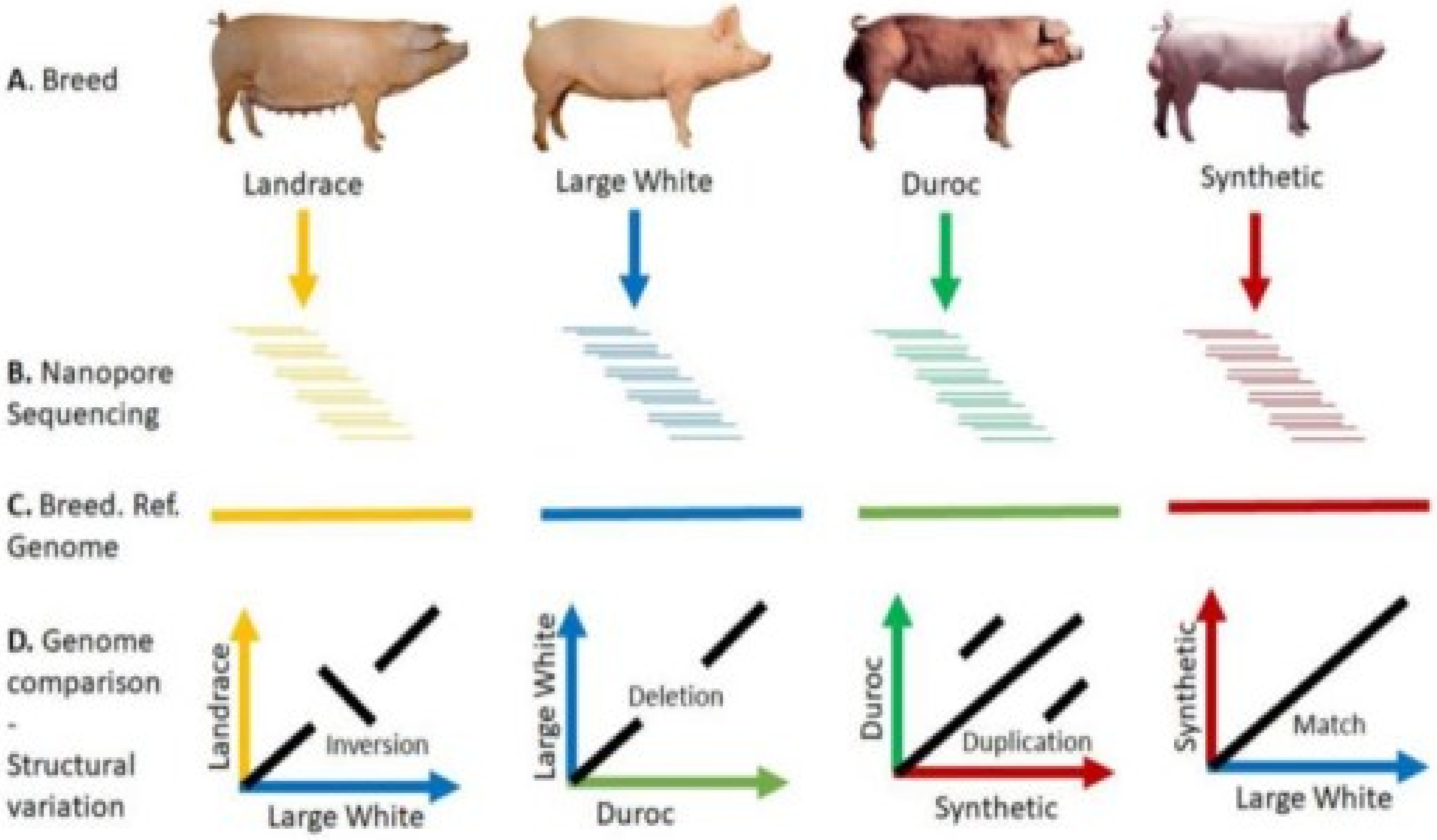
Only GCTA, PLINK gives many false-positives



30 mln SNPs on 25,000 sows

PAN-GENOME WITH LONG READS

DERKS ET AL., 2022



PAN-GENOME

Comprehensive genetic representation of a species, encompassing **all genetic material shared among individuals** as well as the unique genetic variations present in different members of the species.

It can be used as an **alternative or complementary to the reference genome.**

The pan-genome can be particularly useful **to study the structural variation.**

BENEFITS OF PAN-GENOME FOR INSECTS

- **Evaluation of genetic diversity between populations of the same species**
- **Studying the adaptation and evolution of the species under changing environmental conditions – *detecting which regions are conservative***
- **Securing conservation efforts**

World Congress on Genetics Applied to Livestock Production

Madison Wisconsin USA
12-17 July 2026

Sign up for meeting updates!

scan the QR code or visit
tinyurl.com/WCGALP26

SAVE THE DATE

JULY 12-17, 2026

POZNAŃ
UNIVERSITY
OF LIFE SCIENCES

INSECT-IMP

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

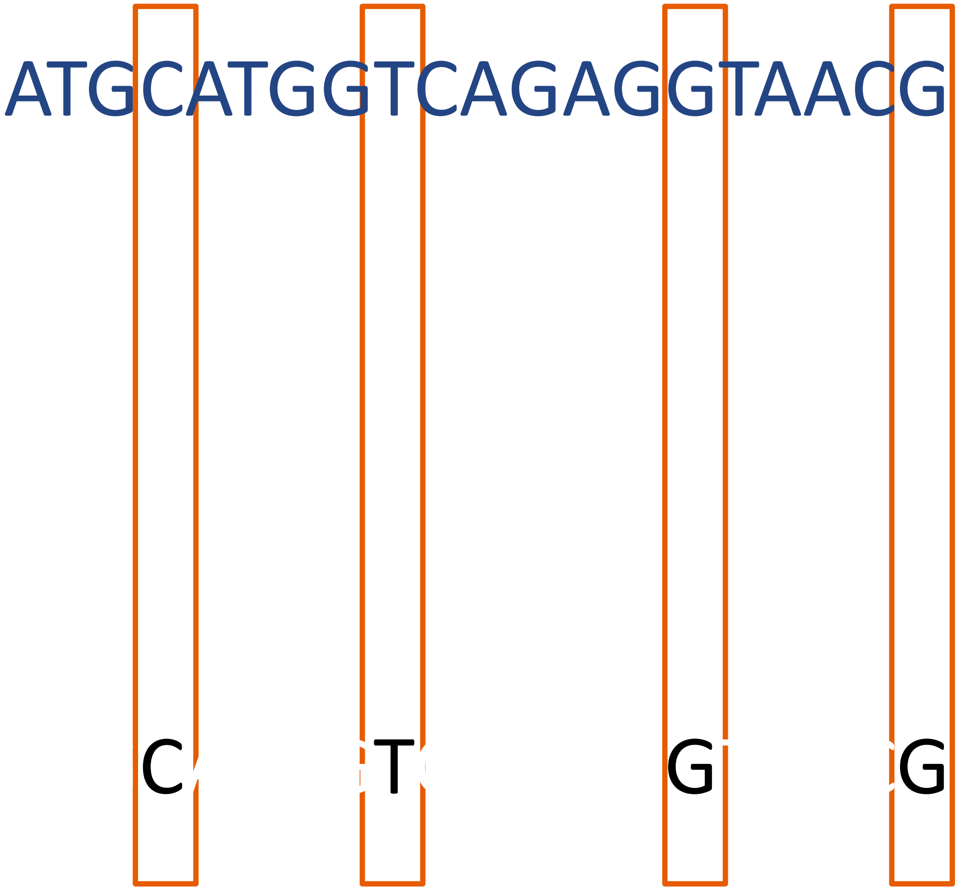
Funded by
the European Union

Thank you for your attention!

Ewa Sell-Kubiak

Department of Genetics and Animal Breeding
Faculty of Veterinary Medicine and Animal Science

IMPUTATION



660k genotype
(High density)

ATGCATGGTCAGAGGTAACG

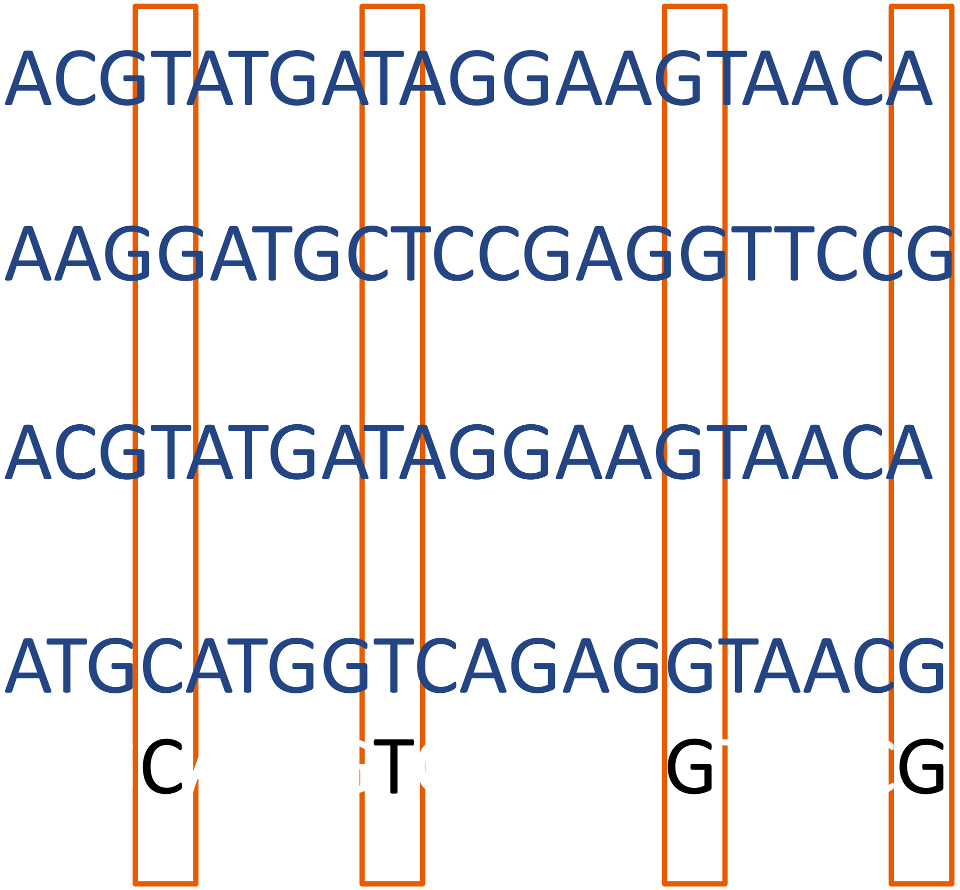
80k genotype
(Low density)

C T G G

by M. Pszczoła

IMPUTATION

660k genotype
(High density)



ACGTATGATAGGAAGTAACA

AAGGATGCTCCGAGGTTCCG

ACGTATGATAGGAAGTAACA

ATGCATGGTCAGAGGTAACG

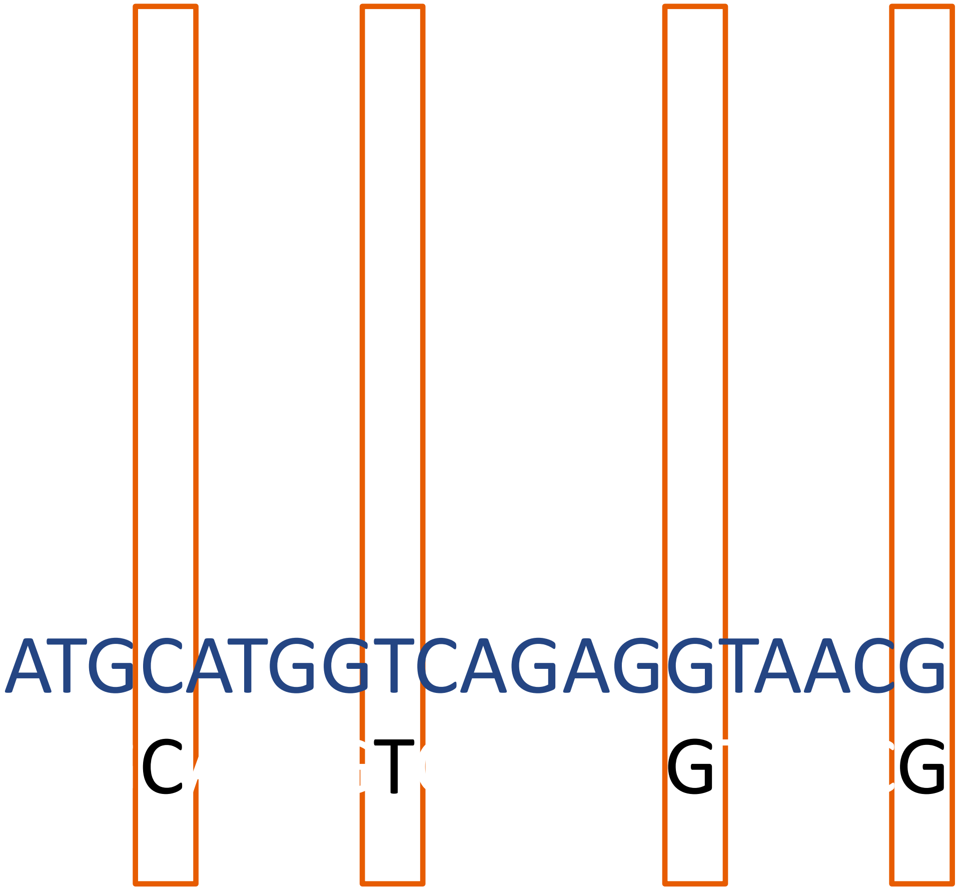
C T G G

80k genotype
(Low density)

by M. Pszczoła

IMPUTATION

660k genotype
(High density)



ATGCATGGTCAGAGGTAACG

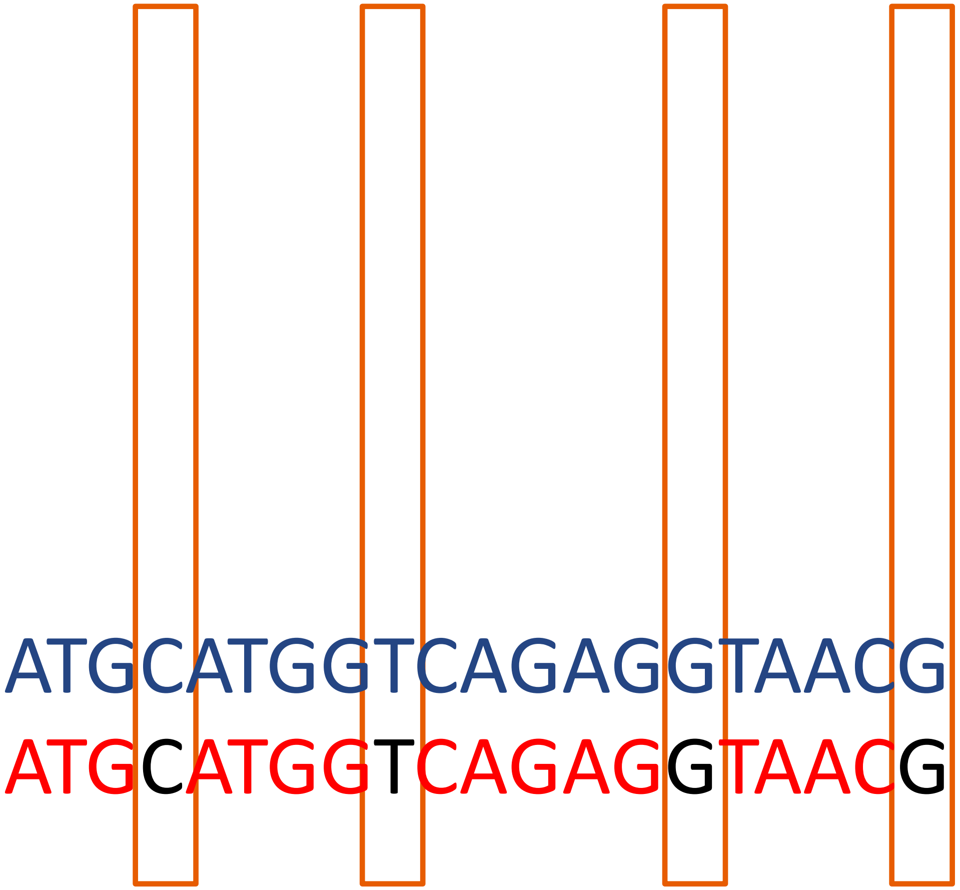
C T G G

80k genotype
(Low density)

by M. Pszczoła

IMPUTATION

660k genotype
(High density)



ATGCATGGTCAGAGGTAACG

ATGCATGGTCAGAGGTAACG

80k genotype
(Low density)

by M. Pszczoła